AD-A111 029   ARIZONA UNIV TUCSON DIGITAL IMAGE ANALYSIS LAB   F/G 5/8

AUTOMATION OF IMAGE PROCESSING.(U)

MAY 81   B R HUNT   N00014-80-C-0344

UNCLASSIFIED   SIE/DIAL-81/007   NL

1 OF 1
AD
A111029

END
DATE
FILMED
8-82
DTIC

No AD
11/30

AD A111029

1

SIE/DIAL 81-007

"AUTOMATION OF IMAGE PROCESSING"

Final Report

ONR Grant No. N-00014-80-C-0344

by

B. R. Hunt

May 15, 1981

Department of Systems Engineering and
Optical Sciences Center
University of Arizona
Tucson, Arizona 85721

ENGINEERING EXPERIMENT STATION
COLLEGE OF ENGINEERING
THE UNIVERSITY OF ARIZONA
TUCSON, ARIZONA 85721

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: SIE/DIAL 81/007

2. GOVT ACCESSION NO.: AD-A[illegible]

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): Automation of Image Processing

5. TYPE OF REPORT & PERIOD COVERED: Final

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): B. R. Hunt

8. CONTRACT OR GRANT NUMBER(s): N00014-80-C-0344

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Systems Engineering Department
University of Arizona
Tucson, Arizona 85721

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:

11. CONTROLLING OFFICE NAME AND ADDRESS:
Information Systems
ONR Arlington, Virginia

12. REPORT DATE: MAY 15, 1981

13. NUMBER OF PAGES:

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report):
Unlimited

APP[illegible] PUBLIC RELEASE
DIS[illegible]D

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):
Unlimited

18. SUPPLEMENTARY NOTES:
None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number):
Image Processing, Automations, Image Analysis, Artificial Intelligence.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

This is the summary report of a study group in the Automation of Image Processing active during 1980, jointly funded by NSF and ONR.

DD FORM 1 JAN 73 1473

SECURITY CLASSIFICATION OF THIS PAGE

FOREWORD

This report is the final documentation of a study-group effort which took place during the year 1980. The purpose of the study-group is appropriately summarized in the title of the original Grant Request: "To Formulate the Automation of Image Processing." But what is image processing and what do we mean by the automation of it?

Image processing has been a very active field of research for the past ten years. Research in academic, governmental, and industrial laboratories has resulted in the successful demonstration of a number of diverse techniques for processing of images. In the following we will define image processing not by a rigorous or a-priori definition, but a-posteriori by example. The basic example of image processing is the human ability to utilize the visual imagery produced by the human visual system. This one example, in fact, is the most sophisticated and complex example we can point to. It is an ability which motivates research in image processing at all levels. Success in machine-emulation of vision is not yet forseeable, however. Much of the successes in image processing, therefore, are associated not with machine vision systems of great flexibility, sophistication, and adaptability, but with machine implementation of processes that are much lower-level in functional contribution to the entire machine vision problem. For example, image processing has been successfully applied to problems such as:

- Image restoration and enhancement [1, 2, 3];
- Image data compression [2, 3, 4];
- Image registration [5,6];
- Stereo-image 3-D model computations [7];
- Synthesis of shaded images [8];
- Extraction of relevant image features, e.g., edges, texture, shapes [3];
- Segmentation of images into disjoint entities [9];
- Use of syntactic and semantic structures to analyze images [10];

This list of techniques is not exhaustive, of course. But even an all-inclusive list would serve to point up the key feature of the field of machine image processing and analysis: That the growing list of techniques still lacks synthesizing principles. That is, there is no overall framework which makes evident how the endless list of specific techniques can be assembled into a system that accomplishes a desired task.

It is this last sentence in the previous paragraph which embodies a functional definition of "automation" for the purposes of this report. The members of the study group adopted a philosophy of defining automation in terms of an external view. That is, imagine an image processing or analysis task of some kind. This imaginary task can be performed by a human being. Automation of that task would constitute the substitution of

a machine which could perform the same task as well* as the human. By taking this external view we are then forced to ask the questions that tell us what must be answered in order to achieve automation, i.e., what is the nature of the data input to the task, how is that data represented, how is the raw input data abstracted to form primitive elements of information, how must the primitives be linked to structure the analysis? This list of questions will be presented implicitly in a model which we discuss in section II of this report.

We can at least give a general answer to the question posed above: what do we mean by the automation of image processing? We mean the machine implementation of processes that perform some desired task, when the major format for input data (but not necessarily the sole or exclusive data format) is an image or images. Using this definition there are clearly some tasks which are approaching automation (e.g., image data compression), and a vast array of other useful tasks which are by no means automated.

Our definition, being a task-oriented one, also admits of quantification in the sense of cost-benefit analysis. Given a specific task it is possible to assess the precise resources, whether machine or human resources, which are required to accomplish the task. The costs of the two sets of resources, human or machine, can then be compared and the issue of whether or not automation would be desirable can be settled.

Footnote: "as well" is a criterion requiring quantification, an issue which the study group recognized as important.

This is the general background of the discussions undertaken by the study group, and is the source of motivations which led to the issues and conclusions presented in the remainder of this report.

Study Group Procedures

The actual procedures by which the study group undertook its work will be briefly summarized in the following paragraphs.

1. Selection of participants. The diverse nature of the proposed study group required that the participants be drawn from as broad a range of image processing interests as possible. Consequently, a list of individuals, along with the professional interests and research expertise of each individual, was compiled. The total number of individuals chosen to participate in the study was deliberately held low, in order to facilitate the informal, open, and uninhibited exchange that might be required from managing meetings of many participants. Individuals were contacted and final selections made on the basis of willingness to participate. The individuals who took part in the deliberations and conclusions of the group were:

   B. R. Hunt
   University of Arizona
   (Group Chairman)

   Marty Tennenbaum
   Fairchild-Schlumberger Research Laboratory

Judith Prewitt
National Institute of Health

Richard P. Kruger
Los Alamos Scientific Laboratory

Robert Haralick
Virginia Polytechnic Institute

King-Sun Fun
Purdue University

Azriel Rosenfeld
University of Maryland

In addition, Laveen Kanal, University of Maryland was invited to participate. However, a variety of scheduling conflicts kept him from attending any meetings.

2. Study group meetings. Three different meetings of the study group were held. The first took place on February 10 and 11, 1981, on the campus of the University of Arizona in Tucson. The second meeting was held on April 28 and 29, 1980, at a hotel in College Park, Maryland (adjacent to the University of Maryland campus). The final meeting, consisted of informal exchanges between the chairman and participants, and two abbreviated meetings on December 1 and 2, 1980, at the International Pattern Recognition conference in Miami, Florida.

The format for the meetings was a very loosely structured debate. The chairman and recording secretary for the group

was B. R. Hunt, who tried to guide discussion into fruitful or unresolved issues and who tried to distill a sense of consensus from the participants. The chairman's notes from the meetings were written up after the meetings and the participants had returned home. Additions, corrections, or deletions for the meeting notes/summaries were invited upon circulation of same to the participants. Any aspect of the meeting summaries not altered in the circulation process were assumed to be agreed to by the group as-a-whole.

3. Final report preparation. This final report was prepared by B. R. Hunt, drawing upon the records of the meetings. Some sections were prepared by other members of the group, e.g., section III was almost totally written by Rosenfeld. A rather conventional disclaimer is desired by the author, Hunt. Namely, the good things in this report are the fruits of the study group participants. The errors or weaknesses are attributable to the author, who may have erred or failed in conveying the insight of the group's participants.

I. Introduction

A. Importance of Automation in vision problems.

Why is the automation of vision problems important? To answer this question requires that we understand the reasons why images are processed in human, non-automated ways.

First, we begin by recognizing that images, whether formed by optical energy, microwave, acoustic, penetrating radiation, etc., are of utility because they represent some portion of the external world. Virtually all of the higher animals have facilities for vision, which can be considered a measure of the evolutionary advantage to possessing a representation of the world external to the organism. Thus, people and animals "process" images because they are useful in the business of living.

Much of our daily human processing of images is relevant only to each of us personally, e.g., we use our eyes in feeding ourselves, driving to work, shopping, etc. As much interest as there may be in someday creating a machine vision system whose capabilities in this arena would be equal to a human being, this is not a primary motivation for the purposes of this study. Humans also use imagery to fulfill purposes which have a social context. For example, the doctor examines x-rays for patient

diagnosis; the civil engineer examines aerial images to plan a highway; the geologist examines satellite imagery to explore for minerals; the astronomer examines star-plates to understand the nature of the Universe. It is the analysis and processing of images in these more specific contexts which is of interest to this study. Our reason for the concentration on such specific applications of imagery is that there is more "leverage". That is, the application of resources to specific problems can yield to solutions for which the cost/benefit comparison between human and machine processing can be quantified, and the investment in automation immediately justified.

What are some specific problems where we believe that automation has the potential for making significant impact? The following, along with examples of same, were obvious to the members of our study group:

1. Processing of large numbers of images.

   Images are being acquired for so many purposes, that it is not clear that there will be enough trained personnel to deal with them. For example, an average hospital will generate approximately 100,000 patient x-rays in one year. All of these must be inspected for patient care, filed, retrieved, refiled, etc. Various federal programs have

mandated the collection and storage of a variety of images. Nuclear reactors must have x-rays of all critical welds, vessel walls, pipes, and support structures. They must be inspected for component integrity and left on-file for the duration of the reactor's operation. Coal-miners are subject to black-lung disease, and every miner must be periodically x-rayed and the x-ray inspected for disease onset. Our earth resources satellites return thousands of images of the earth every month, and the wealth of information in the images about, resources, agriculture, pollution, etc. requires they be visually inspected. Aerial reconnaissance is an established practice in military operations, and every reconnaissance image must be assessed for military significance. New weapons systems have been developed on the basis of sensor technologies such as radar and infrared, and the human utilization of these new classes of imagery is not natural or simple. The list can continue for many pages.

The basic fact that we assert here is that there are many systems and procedures which require the analysis of large masses of imagery. The sensor technology is growing so rapidly that

we will not be able to process the volume of imagery solely by human means. The costs of <u>not</u> processing the imagery can often be established. The benefits of processing the imagery can be established. The costs of computer systems to analyze the imagery can be established once it is known what level of machine function would be required to succeed at a given task. We assert that the cost/benefit analyses of a large number of imagery-related problems will show the desirability of automating these problems.

2. Maximum exploitation of unique images.

At the other extreme from the large volume of imagery is the unique, one-of-a-kind imagery. Imagery from the surface of the moon or from orbiting around a distant planet; imagery from the camera monitoring the hold-up of a bank; imagery from the assassination of a President. Take this last case. Many images of the assassination of John F. Kennedy were taken by amateur and professional photographers the day of the crime. Some of these images were subjected to some of the available techniques for image enhancement. Not all were processed, and not all techniques were tried, because of the limits in time and money imposed by the Congressional investigation.

Questions still remain about the assassination. Does the answer to any such question lie in the unprocessed imagery, or in untried processes?

Given the ubiquity of image sensors, we believe that there will always be images which are so unique that the maximum extraction of information from those images will be an objective. But, as the techniques available for processing and analysis of images continue to grow, the ability of any group of humans to exploit available techniques, and obtain the last iota of information, will be limited. To be able to automate the analysis and processing of such unique cases will be of benefit.

3. Autonomous systems and vehicles.

Our modern world is characterized in many ways, but one of the most significant is the proliferation of vehicles of all kinds. Virtually all vehicles today are under human control, but there are a growing number of cases where human control is not practical or ruled-out. The ability of a vehicle to take antonomous action will require sophistication in a variety of properties referred to as Artificial Intelligence. But it is clear that regardless of the machine intelligence itself, a major input of data for any such vehicle will come in visual

form.

For example, consider the unmanned reconnaissance vehicle, able to explore the enemy's military disposition and either report back or take offensive action on targets of interest. Or consider, the need to assess the interior of a damaged nuclear reactor, such as Three-Mile Island. The unmanned exploration of remote planetary surfaces is an absolute necessity when the time to radio control commands is several minutes or hours. It is estimated that great mineral wealth lies on the bottom of the ocean, inaccessible to manned submarines, but not inaccessible to an autonomous mining vehicle, capable of operating on its own. There are miles of power-lines, pipelines, and railroad tracks which are currently inspected by humans flying or driving past them. An autonomous inspection vehicle could relieve the human burden of this task.

There are autonomous systems that would have uses in non vehicular environments. The classic one is the assembly of complex machinery on a production line, or the test and inspection of the final product. In all such cases, vision and imagery can occupy a critical role in the autonomy of the system.

B. Difficulties in vision automation

It is not the case that the automation of vision and image processing is an untried problem. Many prominent and competent researchers have dealt with this problem, and success is not assured. Results are encouraging, but the date when the successful system is deployed is still not fixed.

The reason that full automation of vision is not at hand is simple: vision automation is a very difficult problem. The problem is even more frustrating in light of how effortlessly human beings can process and analyze the images produced by their visual systems. The real difficulties are surveyed at specific levels in section II of this report, but suffice it to say the following. The state-of-the-art in computer processing is algorithmic, that is, a given process must be expressed in terms of a fixed number of steps, procedures, and decision rules. To make a given algorithm more flexible, the nature of the problem is usually expressed in terms of a model, and the algorithm adjusts some of the parameters of the model in proceeding through the analysis. For the automation of vision, our ability to formulate globally useful models is well-below the requirements demanded by the problem; and our ability to understand the analysis process and formulate all the elements of a global algorithm is equally weak. Thus there is no global model for machine vision and no all-encompassing algorithm. Part

of this study group's contribution to this situation will be seen in subsequent sections.

C. Overview of the report.

In the following sections of the report we will expand upon a model-based view of the image processing and vision problems, and try to formulate an outline of a general-purpose system. We will also survey the state-of-the-art relevant to vision systems and conclude that the state-of-the-art contributes in only a minimal or fragmented way to the envisioned general purpose system. Finally, we will offer recommendations that we believe will facilitate the development of such systems.

II. Model-Based Descriptions of Vision

In myriad areas of computer science, the concept of a model is encountered. The nature of a model is to summarize, in mathematical or descriptive symbolism and language, the basic elements and processes of some existing physical system. The model allows the analyst to investigate the workings of reality within a computer. If a model can mimic the behavior of a real physical system, then the model's utility is twofold. First, the model can be used in a predictive fashion to investigate the way the real-world system would respond under circumstances that have not been previously presented to it. Second, the accuracy of the model in both predictive and analytic modes indicates that the mathematical and symbolical description used in the model is, in some sense, an acceptable abstraction of the fundamental physical mechanisms which govern the working of the real system.

Given the ubiquity of models and modeling processes, it is expected that a desire to model vision processes would be an elemental part of the machine vision problem. The modeling of human vision is made difficult by the inaccessibility of cognitive and physiolocial processes, and by the complexity of these inaccessible processes. The study group felt that in spite of these difficulties, the advancement of our understanding of the vision problem requires that the effort to develop models of vision be undertaken. A great amount of debate in the meetings of the study group centered, therefore, on the strategy by which models of vision could be developed and employed in the analysis of vision. This

discussion of modeling also highlighted another theme: the understanding of vision as a computation process.

A computation process, whether of vision or a numerical algorithm, has clearly defined elements. There are the fundamental units, or atoms, of information which the process operates upon. There are basic processes which are allowed in combining the fundamental units. There are structures which relate elements at one level to elements at a higher or lower level. Finally, there are control mechanisms which the capacity to examine activities at different levels, to initiate or terminate activities at different levels, and to communicate necessary control information between levels.

This structure, of elemental units, levels, and control, was initially employed by the study group as a paradigm of the computational process by which one would describe vision. It was later superseded by a different model, which will be described at the end of this section. First, we wish to discuss the nature of models which will be employed in any paradigm that we use to structure the vision problem.

II.1 Model Descriptions of 3-D World Objects

Vision is a process of perceiving the world external to the viewer. The world external to the viewer is a 3-dimensional world. The objects which reside in that 3-dimensional world have a variety of properties. First, is the diversity of shapes. Objects consist of surfaces, assembled into a variety of shapes. Some shapes are regular, such as the shape of a cube or cylinder. Some shapes are composed of an assemblage of a number of simpler shapes, such as a house being composed of a variety of planes intersecting at different angles. Some shapes are irregular or are made up of such a diversity of surfaces as to be hard to simply describe, e.g., the shape of a fine racing car such as a Ferrari or the shape of something as mundane as a cow.

It is clear that all sighted people are able to deal with the shapes of 3-dimensional objects. That is, we all carry within our heads a variety of "models" about the objects which we have learned to perceive. These models are robust, making it possible to perceive a cube or a cow in a variety of circumstances.

Complicating the problem further is the fact that shapes are not the only mechanism which enters into the process by which vision perceives the world. Objects in the 3-dimensional world possess many other attributes, such as: reflectivity and/or surface roughness; color; being an active source of light or a passive reflection; 3-dimensional solidity and depth; etc. It is clear that our vision system may use some or all of the available attributes in the process of viewing the 3-dimensional world.

To serve in the machine structuring of vision, we must utilize models and descriptions of 3-dimensional world objects. The study group accepted as an article-of-faith, an axiom in fact, that these models and descriptions of 3-dimensional world objects must be dependent in their complexity upon the specific vision application being examined or the goal being pursued. Thus, the modeling of the 3-dimensional world in terms of six-sided regular solids is suitable for simple experiments in the "blocks-world" []. It is not suitable for the autonomous vehicle on the surface of Mars; nor even is it suitable for a robot handler looking for a part on the assembly line.

The models or the descriptions of 3-dimensional world objects should be robust, in the same sense that we know human perception of the visible world is robust. They should be a delicate balance between completeness and parsimonionsness, i.e., between a description of sufficient sophistication to be robust, but without such a plethora of parameters and options as to be unwieldy.

To summarize, the study group felt that substantial work remained to be done in constructing flexible, complete, robust, parsimonious models of 3-dimensional world objects.

II.2 Descriptions of 2-D Scene Creation.

The process of vision, which begins in the 3-dimensional world, progresses through the mechanisms of the vision sensor into the creation of a 2-dimensional scene. That is, energies in the optical energy spectrum which are emitted or reflected by the 3-dimensional world are intercepted by a vision sensor and brought to a focus in the focal plane of the vision sensor. In human vision the sensor is the eye and the focal plane surface is the network of cells and vessels known as the retina.

It is important to note that vision is *inherently* a mapping from the 3-dimensional world into the 2-dimensional world which is at the focal plane of the sensor. Even human vision, which is recognized as possessing 3-dimensional perception, achieves this perception by synthesizing 3-dimensional information from the 2-dimensional scenes registered on the retinas of the left and right eyes. Thus, it is essential to the vision problem that we model the process of the creation of the 2-dimensional scene from the 3-dimensional world.

This problem is much more clearly understood than the modeling and description of 3-dimensional world objects. The processes by which the 2-dimensional scene is created by the sensor is the object of study of two fields which achieved substantial maturity: optics and photogrammetry. Optics is a generic term which we apply to all methods of remotely sensing the 3-dimensional world, and could include image sensors not conventionally associated with classical optics, e.g., imaging by penetrating radiation or with radar waves. Conversely, photogrammetry is the science

which measures and quantifies relations in 3-dimensional world objects from the 2-dimensional scene captured by the sensor.

Optics is mature in describing the physical processes of sensors. The body of two-dimensional linear systems theory, i.e., convolution of optical point-spread-function with the radiance distribution of objects, is adequate to describe the way in which energy emitted or reflected by the object appears in the sensor focal plane. However, this is a 2-dimensional description; that is, even in the case where the optical point-spread-function induces no degradation (point-spread-function is a Dirac function) the image appearing at the focal plane surface is a projection intc 2-dimensions of the 3-dimensional world. Understanding the nature of this projection is the subject of the science of photogrammetry. From knowledge of key parameters in the sensor, e.g., focal length, size of image on the focal plane surface, operating f-number, camera orientation and position in 3-dimensional space, etc., it is possible to model the geometry of the process of projection from the 3-dimensional world into the 2-dimensional scene.

A final element in the scene modeling process is the image detector. It must be emphasized that scene creation is a dynamic process. A scene is formed in the sensor focal plane by the propagation of energies from the object to the sensor focal plane. But an image is a record of the energy distribution at some point in time. This is true whether the sensor is the eye or a camera. Therefore, it is the function of an image focal plane detector to capture the transient energy flow and make a record of it at some point in time. It must be noted that this record will be imperfect.

Every detector introduces noise in the process of detecting and recording the image. Thus, the image at hand is the original transient flow as made available through the detection mechanism.

Again, the phenomenology and physics of image detectors is mature, when compared to other parts of the vision problem. The specific operating parameters of a detector such as the human eye may be unknown, but the general mathematical descriptions are adequate for most applications.

The actual nature of the models to be employed for describing the creation of the 2-dimensional scene from the 3-dimensional is dependent upon the specific sensor and its scene creation mechanism. For example, some commonality exists between optical photography scene creation and x-ray scene creations (see, for example, Andrews and Hunt) but the commonality must be augmented with specifics that make the differences recognizable.

The study group felt that mechanisms of scene creation are well-known enough so that further research in this area is not of high priority. However, one area in which research is warrented is the marriage of descriptions of 2-dimensional scenes formed from models of 3-dimensional world objects. That is, it would be useful to have compact, simple, and easily used mechanisms which would take a particular set of models of 3-dimensional world objects, impose the sensor processing, account for the projection operations between 3-dimensional and 2-dimensional, and create a 2-dimensional focal plane scene. Parts of this problem are solved (e.g., the hidden-line problem for simple graphics) and parts are under constant attack (e.g., the image animation and scene synthesis

research in 3-dimensional computer graphics). Whether the solutions known are relevant to the automation for vision problems is not clear, however.

II.3 Inference of 3-D Objects from 2-D Scenes

The most difficult part of the modeling is the process by which vision infers the nature of the 3-dimensional world from the 2-dimensional scene. That such occurs is self-evident, and that such seems to be an effortless process is also self-evident. What mechanisms can we suggest to achieve this inference process. There are three which the study group found relevant.

1. The Optimization-of-parameters approach.

   In this case the inference process is conceived of as something similar to the fitting of a model, as in statistics by least-squares. For example, if a model for the 3-dimensional world exists which possesses a number of unspecified parameters, then changes in those parameters will change the actual way in which the 3-dimensional world is being presented. For a given fixed set of values for all the parameters it is possible to take the corresponding 3-dimensional world model, pass the model through the equations describing the sensor and its projection geometry, and construct the representation of the 2-dimensional focal plane scene which corresponds to the specific set of parameters in the 3-dimensional world models. The difference between the 2-dimensional scene generated

by this synthesis process, and the actual 2-dimensional scene observed by the sensor, is attributable to either:

a. differences in how completely the 3-dimensional model describes the world being imaged;

or,

b. if the model is complete, a set of model parameters which is at variance with the actual description of the 3-dimensional world objects.

Assuming that one could define a metric between the scene synthesized through the sensor and the actual recorded scene, it would then be possible to conceive of varying the parameters of the 3-dimensional world model until this metric is minimized. This optimization approach is conceptually straight-forward; it is based upon a science which is relatively mature, i.e., optimization theory. In reality, however, the details for a successful implementation remain very complex. Besides the modeling problems described above, there remain also the establishment of a metric between the actual scene and the scene representation

for a specific set of parameters. Following the establishment of a metric, it would be necessary to determine how the minimization of a metric could be related to the parameters which are being varied. Thus, the optimization approach remains conceptually possible, but computationally difficult.

2. The Heuristic Approach

Lacking the rigor of an optimization approach, the other extreme is the inference of the 3-dimensional world from the 2-dimensional scene as a purely heuristic approach. Here the specifics of the individual problem would be used to guide the development of an algorithm which would use the 2-dimensional scene to guide the structuring of the 3-dimensional world. This is a specific approach, since the algorithm applied to one particular 3-dimensional world would probably be unworkable or errorneous for small or moderate changes in the 3-dimensional world. (By contrast, the optimization-of-parameters approach is general; its difficulty lies in its generality, in fact.)

Much of the work which has been done in vision to this point could be fairly described as lying much closer to the heuristic approach than to an optimization model approach. The difficulty of the general optimization approach is one reason for this. Another reason is that what knowledge we have of how human vision works is derived from

psychological experimentation that lends itself to being described in various types of heuristics. Thus, heuristic algorighms for the inference of the 3-dimensional world from the 2-dimensional scene also represent an approach to human vision. Finally, the heuristic approach is attractive because of its pragmatism. If we are confronted with a specific problem, why not concentrate on developing methods that are directed to just that problem and no others? Such an approach can be developed <u>independently</u> of any heuristics motivating human vision.

3. The Human Vision-System Approach.

If we understood the mechanisms of human vision, the obvious consideration would be to create a system which operated on the same principles. Although obvious, this is by no means reasonable. For example, we might be confronted with a firm knowledge of how visual processing migrated from the retina to the cortex and beyond, and be forced to acknowledge that such a mechanism involved unique properties of biological systems that could be replicated in machines with the requisite quantity only at impossible cost. Stated another way, the understanding of human vision would not imply that such understanding could be replicated economically in machines with the flexibility of human vision.

The cost-effective approach to the vision system with the greatest flexibility may remain as the human being for many years to come.

Nonetheless, the study group felt that an understanding of the human vision system would offer significant insights into the structuring of general-purpose vision systems.

II.4 A General Paradigm for Vision Systems

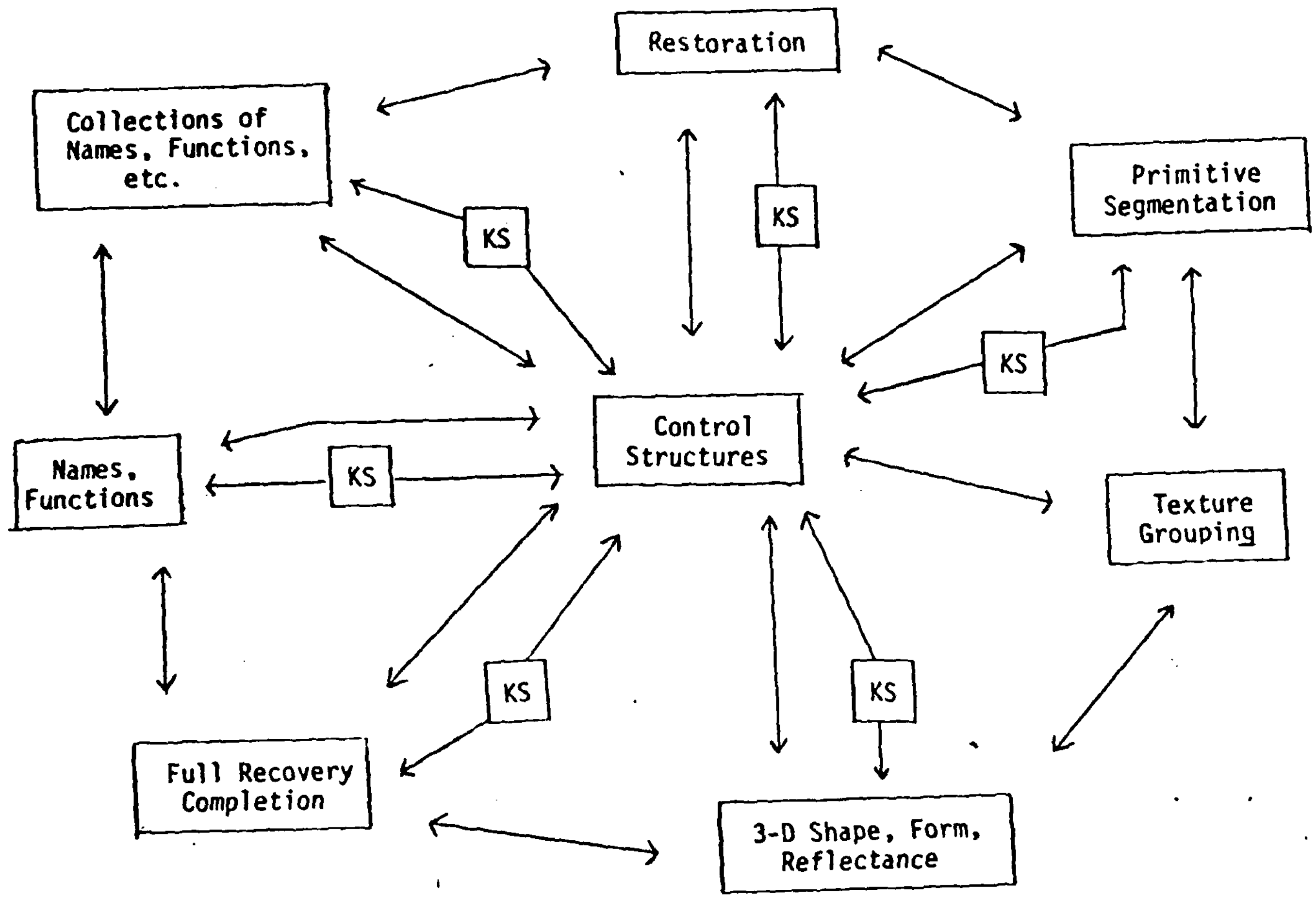
The above paragraphs present the study group's views of the necessity and utility of models for the vision process. But the models must be used within a structure. And in the following we present a paradigm for the usage of the models, a paradigm for vision systems of any kind. It is into the framework of this paradigm that any of the above discussion can be placed.

We begin the discussion by noting that there is a temptation to think of the problem of constructing a vision system paradigm in terms of levels; that is, to perceive the problem, followed by the next level, etc. In passing through such levels the progression into the problem is one of simpler levels in the upper portions of the problem, with each succeeding lower level being richer in complexity and structure. The temptation is a natural consequence of modern techniques in computational algorithm analysis: the top-down analysis of a problem. In its initial deliberations the study group adopted such a top-down structure. However, in thinking about applying the top-down structure to any of a number of applications the study group found the top-down,

multiple-layers approach was continually being circumvented by various qualifying statements, e.g., for a given problem it was observed that a specific level of the paradigm did not exist for this particular problem, or that the necessity existed to allow levels to be bypassed for this problem.

A further difficulty was the question of control structures, that is, the mechanisms which mediate the transfer of data between problem levels, or which activate computational and/or analysis processes within the levels. It is evident that control structures span all the levels, but how to best represent this? Finally, the levels approach also was weak in recognizing the way problems do not pass uniformly from level to level. That is, the paradigm of top-down levels leads to an orientation in which analysis passes from level to level. Whereas, actual processing may require re-tracing levels, that is, processing taking place at level N may suddenly reach a dead-end and require a restart at level N-1 or level N-2, N-3, etc.

The difficulties with a top-down levels structure was recognized as one more of conceptual organization and topology than an improper analysis of the vision system problem. The study group resolved it by structuring the vision system paradigm not with multiple levels but as a ring. This paradigm retains the necessity for separate functions, but does not force them into the strait-jacket of levels. Finally, the inclusion of knowledge sources/ data bases appropriate to each problem function is simple in this paradigm. The resulting structure is seen in the accompanying drawing, Figure 1. The ring is made of the elementary functions



where KS = Knowledge Sources

Figure 1

of the vision system, centered about the control structures; the centrality of control structures makes it possible for control processes to draw upon the knowledge sources associated with level, and to pass information between separate functions. The specific functions which are present in the paradigm are as follows:

1. Restoration.

   This is the process of employing models of the sensor and image formation process so as to invert or remove any deleterious effects of image formation and provide an image that corresponds as closely as possible to the original object radiance distribution.

2. Primitive Segmentation.

   Understanding of the image requires that different entities— segments— be recognized as existing in the scene. This problem has received much attention in research. The basic methods popular in research rely upon the computation of primitive properites of the image, e.g., edges, gradients, color clusters, etc., and to use similarity measures (or dissimilarity measures) to group sets of pixels together into contiguous segments.

3. Texture Grouping.

   Conceptually a part of the wider class of functions we referred to as segmentation, texture is by itself such a rich a varied key to image segments as to merit a separate function block. It is also one of the most difficult problems; textures that are subtly different, yet easily distinguished by the human viewer,

are not easily distinguished by machine.

4. 3-D Shape, Form, Reflectance.

   As noted previously, the image is a mapping from the 3-dimensional world to the 2-dimensional scene. One of the most important tasks, therefore, is to recover intrinsic scene characteristics. The first level of this recovery problem is the 3-dimensional shape and form of an object. Reflectance, and reflectance discontinuities, are also an important part of the shape and form recovery process.

5. Completion of Full Recovery.

   When completed, not only shape, form, and reflectance are understood, but also the host of other cues which a vision system can use: stereo depth/distance, object orientation, relative location of objects with respect to each other, etc.

6. Names, Functions.

   Once object recovery is complete, the labeling of objects can take place (assuming that recovery not only distinguishes objects, but furnishes the information for unique names or functions to be associated with the distinguished objects).

7. Collections of Names, Functions.

   Recognizing that vision can fruitfully organize objects into hierarchies, the names and functions can also be so organized.

The completion of the ring, showing links between all functions, as well as links to the center, is part of the strategy, discussed earlier, to outline the problem functions of a vision system without being bound by the top-down level concepts that are the first reaction to an anlysis of the vision problem.

The study group does not believe that the vision system functions identified in Figure 1 are anything new in the description of vision; other workers have identified similar functions. The contribution of the study group is in the ring structure, with central coordination by control structures, accessibility (through control) of functions or knowledge bases between problem functions, and the flexibility of allowing control to activate appropriate functions without following a strict sequencing from function to function around the ring.

III. The State of Research Relevant To Vision Systems

The state of the art in various applications of image processing and computer vision will be discussed in the next section. This section briefly deals with general aspects of the subject, with reference to the processes referred to in the paradigm described earlier in this report. Only selected major research themes are mentioned; much work needs to be done on a wide variety of specific techniques, in order to improve their performance and establish their effectiveness.

III.1 Restoration

Image restoration techniques rely upon the most simple of sensor models, space-invariant image formation with signal independent noise. These models are adopted primarily for computational tractability, since these models can be implemented through fast Fourier algorithms. Realistic sensor models, however, must include space-variant image formation so as to accurately model a wider variety of image formation processes, e.g.: space-variant imaging for accelerated motion imagery; signal dependent noise for photon-statistic noise; nonlinear sensors, with imaging kernels dependent upon either object structure or object brightness level; nonlinear recording phenomena.

Image restoration processes also suffer from a lack of robustness, i.e., they all require a minimal amount of input information in order to function. Some methods have been developed for the problem of "blind deconvolution", image restoration in the absence of the usual minimal input information. A certain

amount of success has been achieved with blind deconvolution, i.e., enough success to demonstrate the validity of the objective, as well as demonstrate how remote the final goal truly is.

Finally, restoration could be described as signal-processing oriented, i.e., based upon image-in, image-out processes with common signal processing functions carried out, processes such as convolution, correlation, Fourier transform, etc. The ring paradigm of the previous section indicates how restoration can be linked to analytical processes that are not of a signal processing nature. One can conceive of schemes in which restoration is linked to the processes of recovery, grouping, etc.

III.2 Segmentation

Most segmentation techniques involve classifying the image's pixels on the basis of gray level (e.g., light vs. dark), spectral signature, or local property values (e.g., high vs. low gradient magnitude, or degree of match to some local template). If these classifications are done independently, they can be implemented on parallel hardware, so that the segmentation process becomes very fast. Alternatively, they can be done sequentially (e.g., region growing), and the classification criteria can be progressively refined as the segmentation proceeds, so that context, as well as knowledge about global and geometric properties of the segments can be used to guide the segmentation process; but sequential processes are inherently slow.

Some recent work using probablistic classification and multiple-resolution image representations may provide a basis for defining segmentation techniques that are spatially parallel but that allow contextual and global information to be incorporated. "Classifying" each pixel probabilistically, and iteratively adjusting the probabilities based on those at the neighbors ("relaxation") allows the classification to be refined based on the context. Analyzing a "pyramid" of images at successively lower resolutions allows certain types of global information to influence the process (e.g., regions of various shapes become local features of particular types higher in the pyramid, and the presence of these can bias the processes at the lower levels).

Another approach to segmentation is based on partitioning the image into maximal homogeneous regions. Efficient split-and-merge algorithms for constructing such piecewise approximations have been developed. A piecewise constant model is most commonly used, but better results can often be obtained using piecewise linear approximations. Mathematical methods of finding optimal piecewise approximations are a subject of active research.

Any segmentation technique implicitly assumes a model for the class of images on which it will be effective. The models assumed by the standard techniques are highly oversimplified (e.g., simple assumptions about the gray level population of the image, without regard for spatial arrangement), but models

incorporating contextual and geometrical information are gradually being developed. In specific application areas, it would be appropriate to develop specialized models for the particular classes of images being analyzed, and to design optimal segmentation methods based on these models.

III.3 Texture

A variety of methods exist for classifying uniformly textured regions, but it is much harder to segment a given image into such regions. Work is being done on the extension of pyramid-based segmentation methods to the texture domain; here texture differences are reliably detected by comparing large image blocks, and the boundaries of the regions are accurately located by examining smaller blocks. The most commonly used texture features are based on first- and second-order statistics of gray levels or local property values; another possibility is to compute such statistics for properties of "primitives" extracted from the texture.

III.4 Disambiguation

Recovery of 3-dimensional scene information from an image or set of images is an active area of research. The possible (or plausible) 3-dimensional configuration corresponding to a portion of an image is constrained by shading, texture gradients, nature of edges present, and 2-dimensional contour shapes; methods of

ference between the 2-dimensional scene generated

quantifying and using these constraints are being developed. Such methods are also being extended to stereopairs and to time sequences of images ("optical flow").

The ideal goal of segmentation is a decomposition of the image into objects or surface patches, not into two-dimensional regions. Given the "intrinsic images" (range, slope, illumination, reflectivity, etc.) resulting from disambiguation, segmentation becomes much easier; but the disambiguation process itself requires an initial segmentation. The processes of disambiguation and segmentation need to be closely interfaced.

III.5 Shape and Structure

Even in two dimensions, shape description is a nontrivial task. Hierarchical representations (quadtrees, striptrees) appear to be useful in handling shape at coarse and fine levels, but do not directly represent significant parts and their relationships, which are better handled using structural descriptions based on approximations of the parts, e.g., by generalized "cones". Analogous remarks apply in three dimensions.

Scenes consist of related collections of objects, just as objects are composed of parts. Thus hierarchical graphs, labelled with property and relation values, provide a natural medium for representing scene structure, and also for representing models that define classes of scenes. The process of matching scene descriptions to models can be regarded as a form of graph parsing, involving a hierarchy of subgraph matchings. The combinatorics of this matching process can be alleviated by

using "smart search" techniques, relaxation methods, or a combination of both.

III.6 Control

As we have seen, the analysis of images involves many different representations for both the data derived from the image (intensity arrays, region geometry, relational structures) and the knowledge or models used in the analysis, and involves complex interactions among the processes that operate on this information. The control of these interactions is a very difficult task. No theoretical basis as yet exists for designing representations, models, or control structures appropriate for a given image analysis domain. As experience with operational systems in a variety of applications accumulates, theoretical foundations for various aspects of the process are beginning to emerge, so that computer vision is gradually being transformed from an art into a science.

III.7 Conclusions

It is evident from the discussion in the previous sections that there are many open problems. If we use the paradigm in section II (recall Figure 1), then there is not a single "box" where much research does not remain. This is true even though some of the functions shown in the boxes of Figure 1 may be thought to be essentially solved problems; for example, image restoration might be thought to be solved, given the success of research in this area, but such is not true. Even image restora-

restoration is an area in which realistic models and sensors have not been adequately dealt with.

It is the conclusion of the study group that:

1. The creation of a successful machine vision system requires continued and extensive research in all facets of the vision problem;
2. the paradigm of Figure 1 represents a new tool to utilize in structuring both the research in individual functions of the vision system problem and in the coordination and control of separate functions in an integrated autonomous system.

The study group felt that one danger in real progress toward a vision system is the fragmentation that is taking place in the community of individuals working in all facets of image processing. It is hoped that the paradigm and structure of Figure 1 can provide a framework within which all segments of the vision/image processing community can interpret their work, and relate to other workers and areas.

IV. Other Things Recommended to Facilitate Research

IV.1 Standardized Tools and Data Bases

Large numbers of groups in universities, industry, and government laboratories are working on many different aspects of image processing and computer vision, but there is very little exchange of software among these groups, even when they are working on common problems. Exchange would be desirable to reduce duplication of effort, but is difficult because of the complete lack of software standardization; a great variety of machines, operating systems, and languages are being used. A transportable Fortran-based image processing software package is being developed jointly by the University of Maryland and Virginia Polytechnic Institute under NSF sponsorship. This package should be useful for groups entering the field who need to quickly build up an experimental capability; but it is less likely to be used by major research groups or in situations involving production-run processing. Efforts are also being made to achieve some degree of software compatibility on the DARPA Image Understanding Program, based on the use of a small set of common machines, operating systems, and languages; it remains to be seen how effective these efforts will be.

The situation with regard to data bases and data exchange is somewhat more satisfactory. Several standard formats for image data have been proposed, and one of these, the NATO format, is widely used. A number of groups have available for dissemination data bases of various types of digital imagery,

including alphanumeric characters, biomedical images, reconnaissance and cartographic imagery; and in some cases (particularly satellite remote sensing), imagery is routinely archived and made available in digital form.

IV.2 Clearinghouses

The great proliferation of work in this field makes it very difficult to maintain awareness of ongoing activities, available techniques, etc. Research results are presented at a large number of meetings (several per year in each major application area, plus several of a general nature) and are published in a large number of journals. Because of the wide variety of applications, many different professional societies and government agencies are involved, and there is no single focal point for coverage of more than a fraction of the field. It would be desirable to establish some sort of clearinghouse for information on image processing and computer vision, including projects, software, etc., under interagency sponsorship. Regular communication among the agencies who support work in the field would also be very desirable, to promote coordination of efforts and funding.

IV.3 Vision Research Council

Since progress in image processing and computer vision is very rapid, monitoring of the state of the art and of research needs should be an ongoing process. It would be desirable to establish a Vision Research Council composed of authorities in

the field, to carry on such a monitoring process and make recommendations periodically to the cognizant government agencies. If an interagency committee were established in the field, the Council could serve as an advisory body to that committee.

Iv.4 Vision Institute

One way to implement the proposed clearinghouse and standardization activities is to establish a National Vision Institute which would be responsible for these efforts. This Institute could also serve as a base of operations for the Council and provide supporting services (clerical, etc.) for its activities. It could be supported by interagency funding, and possibly also by contributions from cognizant professional societies or industrial organizations.

Defense Documentation Center Cameron Station Alexandria, VA 22314	12 copies
Office of Naval Research Arlington, VA 22217	
Information Systems Program (437)	2 copies
Code 200	1 copy
Code 455	1 copy
Code 458	1 copy
Office of Naval Research Eastern/Central Regional Office Bldg 114, Section D 666 Summer Street Boston, MA 02210	1 copy
Office of Naval Research Branch Office, Chicago 536 South Clark Street Chicago, IL 60605	1 copy
Office of Naval Research Western Regional Office 1030 East Green Street Pasadena, CA 91106	1 copy
Naval Research Laboratory Technical Information Division, Code 2627 Washington, D.C. 20375	6 copies
Dr. A. L. Slafkosky Scientific Advisor Commandant of the Marine Corps (Code RD-1) Washington, D.C. 20380	1 copy
Naval Ocean Systems Center Advanced Software Technology Division Code 5200 San Diego, CA 92152	1 copy
Mr. E. H. Gleissner Naval Ship Research & Development Center Computation and Mathematics Department Bethesda, MD 20084	1 copy
Captain Grace M. Hopper (008) Naval Data Automation Command Washington Navy Yard Building 166 Washington, D.C. 20374	1 copy

FILMED

3-8